

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

# Concise Encyclopedia Biochemistry

Second Edition

Revised and expanded by  
Thomas Scott and Mary Eagleson



Walter de Gruyter  
Berlin · New York 1988

*Concise Encyclopedia of Biochemistry*

first published in 1983 based on the original, German language edition,  
Brockhaus ABC Biochemie

Copyright © 1976; 1981 VEB F. A. Brockhaus Verlag Leipzig

Translated into English, revised and enlarged by Thomas Scott and Mary Brewer

First Printing, March 1983

Second Printing with corrections, October 1983

Second edition 1988 by

Thomas Scott, Ph. D.  
Department of Biochemistry  
University of Leeds  
Leeds, England

Mary Eagleson, Ph. D.  
Berkshire Road  
Sandy Hook  
Connecticut 06482, USA

*CIP-Titelaufnahme der Deutschen Bibliothek*

**Concise encyclopedia biochemistry** / rev. and expanded by Thomas Scott and Mary Eagleson. [Transl. into English, rev. and enl. by Thomas Scott and Mary Brewer]. - 2. ed. - Berlin ; New York : de Gruyter, 1988  
Einheitssach.: Brockhaus-ABC Biochemie (engl.)  
Früher u. d. T.: Concise encyclopedia of biochemistry  
ISBN 3-11-011625-1  
NE: Scott, Thomas [Bearb.]; EST

*Library of Congress Cataloging in Publication Data*

Brockhaus ABC Biochemie. English.

Concise encyclopedia biochemistry. - English language ed., 2nd ed. / revised and expanded by Thomas Scott and Mary Eagleson.

Translation of: Brockhaus ABC Biochemie. ©1981.

Rev. ed. of: Concise encyclopedia of biochemistry. 1983.

ISBN 0-89925-457-8 (U.S.)

1. Biochemistry-Dictionaries. I. Scott, Thomas, 1935- .

II. Eagleson, Mary, 1945- . III. Concise encyclopedia of biochemistry. IV. Title.

[DNLM: 1. Biochemistry-dictionaries. QU 13 B864]

QD415.A25B713 1988

574.19'2'0321-dc19

DNLM/DLC

for Library of Congress

English language edition

Copyright © 1983; 1988 by Walter de Gruyter & Co., Berlin 30. - All rights reserved, including those of translation into foreign languages. No part of this book may be reproduced in any form - by photoprint, microfilm or any other means - nor transmitted nor translated into a machine language without written permission from the publisher. - Typesetting and Printing: Appl, Wemding. - Binding: Dieter Mikolai, Berlin. - Cover Design: Hansbernd Lindemann, Berlin. - Printed in Germany.

The detailed and complete reaction sequence was finally elucidated by the use of further mutants (blocked at various stages between glutamate and ornithine), by the additional use of isotopic methods, and by the characterization of the individual enzymes.

The site of the metabolic block can be further ascertained by measurement of the appropriate enzyme activity. This will be absent or greatly decreased, compared with that in the wild type organism. Sometimes, serologically similar, but enzymically inactive protein is produced, i.e. the mutation has not prevented transcription and subsequent translation of a protein, but small changes (sometimes the replacement of a single amino acid) have occurred which destroy catalytic activity. Such proteins are known as CRiM proteins (Cross Reacting Material; the "i" is added for pronunciation purposes).

Elucidation of the biosynthetic pathway of the aromatic amino acids (see Aromatic biosynthesis) was greatly aided by the use of mutants of *Escherichia coli* and *Aerobacter aerogenes*. Many bacterial mutants were known to have a simultaneous requirement for phenylalanine, tyrosine, tryptophan, *p*-aminobenzoic acid and hydroxybenzoic acid, thus indicating a common biosynthetic pathway for all these compounds. This multiple requirement was satisfied by the single compound, shikimic acid, which can be isolated from the leaves of various plants (Gymnosperms, *Illiaceae*, etc.). Shikimic acid was later found to be excreted by certain auxotrophs with multiple aromatic growth requirements. Subsequently, auxotrophic mutants for each step of aromatic biosynthesis were isolated, the accumulated material was identified and shown to support the growth of mutants with earlier blocks, and the appropriate enzymes isolated and studied. The key branch point compound, chorismic acid, was identified with the aid of a triple mutant, with metabolic blocks in the conversion of chorismic acid into prephenic acid, prephenic acid into *p*-hydroxyphenylpyruvic acid, and anthranilic acid into indole 3-glycerol phosphate. In the presence of L-tryptophan (to repress the formation of enzymes converting chorismic acid into anthranilic acid), washed suspensions of this mutant (*A. aerogenes* 62 - 1) excreted chorismic acid.

The use of mutants, and in particular the supplementation test, are difficult when a single mutation of a fundamental metabolic reaction gives rise to multiple growth requirements, i.e. polyauxotrophy. This must be clearly distinguished from polyauxotrophy resulting from polygenic mutations, i.e. single mutations in several distinct pathways.

The mutant technique has also been applied to naturally occurring mutations in animals, e.g. in the study of the degradation of phenylalanine and tyrosine (see Phenylalanine).

**Mutarotation:** a change in the optical rotation of an optical isomer, usually a carbohydrate, in aqueous solution. The carbohydrate molecule can exist in anomeric forms, designated  $\alpha$  and  $\beta$ . These two diastereomers differ in chemical and physical behavior, such as melting point, solubility, and especially optical activity. In aqueous solution an equilibrium is gradually established between the two

diastereomeric half acetal forms and the open chain form, with a consequent change in the optical rotation. Interconversion of the two diastereomers occurs via the intermediate cyclic half acetal. The attainment of equilibrium is accelerated by acids or bases. The starting value of  $[\alpha]_D^{20}$  for  $\alpha$ -D-glucose in water is  $+113^\circ$ , that of  $\beta$ -D-glucose is  $+19.7^\circ$ . After a few hours, when equilibrium between  $\alpha$  and  $\beta$  forms is established, the value  $[\alpha]_D^{20}$  is  $+52.3^\circ$ , representing a mixture of 37%  $\alpha$  and 63%  $\beta$ -glucose. A monosaccharide showing M. is often characterized by measurement of the starting and final values for rotation, e.g.  $[\alpha]_D + 113^\circ \rightarrow 52^\circ$ .

**Mutation:** chemical or physical changes in the genetic material of a cell or organism. A single M. represents a change in a gene, which is a defined segment of DNA (or RNA). Chemical changes in the DNA include substitution of one nucleotide for another, due to an error in copying (base pairing) or to a change such as dimerization of adjacent bases which prevents accurate copying. Physical changes include breakage and loss of part of a DNA molecule, or rearrangement of the molecule.

A point mutation affects a single nucleotide; it may consist of a) substitution of a different nucleotide, b) loss of a nucleotide (deletion) or c) insertion of an extra nucleotide. In each case the nucleotide sequence is altered; in b and c, the "reading frame", or division of the sequence into triplets to be translated into amino acid (see Genetic code; Protein biosynthesis) is shifted, with severe effects on the protein product of the gene if it is a structural gene.

The substitution of one base for another may or may not change the interpretation of a base triplet or a non-translated stretch of DNA. M. which have no phenotypic effects are called "silent M." A large majority of the DNA in a multicellular organism appears never to be translated. Some portions of this untranslated DNA regulate the expression of structural genes (see Gene expression), some apparently serve to make rearrangements such as crossing over of sister chromatids and movement of Transposons (see) possible, some may serve to promote pairing of homologous chromosomes at meiosis, and others may have no function at all or serve merely as spacers between genes (see Introns). M. in such regions are likely never to be detected.

M. which can be detected result in changes in structural or regulatory genes, and thus are reflected by absence or change in enzyme or structural proteins, metabolic reactions or overall performance of a metabolic pathway. At the morphological level, the results may be, e.g., changes in pigmentation, body structure or response to environmental changes.

M. may be random, spontaneous or induced, and in a multicellular organism, they may occur either in germ-line or somatic cells. (Naturally only the former are passed on to subsequent generations.) A random M. may be selected by environmental pressure for preservation, although it is far more likely to be deleterious. Spontaneous M. occurs because DNA replication is subject to occasional errors, and for reasons which are not understood, some stretches of DNA are more subject to such errors than others ("hot spots"). Genetic variability within a species confers the advantage of adaptability to

## Mutational frequency

changing environmental conditions, and thus evolution appears to have selected DNA polymerases with a low but finite tendency to err in replication. Induced M. is caused by Mutagens (see), which are defined as agents which increase the low rate of spontaneous M.

**Mutational frequency:** see Mutants.

**Mutation rate:** see Mutants.

**MWC model:** see Cooperativity model.

**Mycobactin:** a Siderochrome (see) synthesized by *Mycobacterium* spp.

**Mycocerotic acids:** see Fatty acid biosynthesis.

**Mycorrhiza:** symbiotic association between the roots of a higher plant (forest trees, orchids, etc.) and a fungus. The fungal mycelium covers the root tips, and the root hairs become reduced or disappear. The fungal hyphae take up nutrients from the soil. The hyphae may penetrate the intercellular spaces of the root (ectotrophic M.), or actually penetrate the plant cells (endotrophic M.). Endotrophic M. is necessary for the germination of orchid seeds. Ectotrophic M. are formed by many fungi with forest trees.

**Mycosterols:** see Sterols.

**Mycotoxins:** metabolic products of certain fungi and other microorganisms, which are harmful to other organisms, especially vertebrates, including man. The same M. may be produced by more than one fungal species. Out of about 100 000 described species of fungi, 50 are known to produce M.; these may damage the host directly (e.g. plant pathogenic fungi), or indirectly by causing illnesses in animals and man when M. are consumed in the diet. Ergotism is a classical example of mycotoxicosis (see Ergot alkaloids). M.-producing organisms frequently develop on improperly stored foodstuffs, leading to food poisoning. Such M. are e.g. botulinus toxins (see Toxic proteins), Aflatoxins (see) and Ochrotoxins (see). Other M. include the nephritic toxin citrinin from *Penicillium citrinum*, notatin from *Penicillium notatum* and sporidesmin from *Pithomyces chartarum* (formerly *Sporidesmin bakeri*). Other important M. producers are *Penicillium islandicum*, *Penicillium rubrum*, *Paecilomyces varioti*, *Fusarium sporotrichioides* and *Stachybotrys atra*. M. also include the bacterial toxins, which are subdivided into endo- and exotoxins (see Toxic proteins).

**Mydriatic alkaloids:** see Tropane alkaloids.

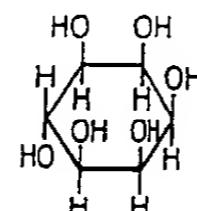
**Myelin protein A1:** see Encephalogenic protein.

**Myoglobin:** a single chain heme protein of skeletal muscle;  $M_r$  17 200; 153 amino acid residues. Function of M. is oxygen storage and transfer (i.e. from hemoglobin to respiratory enzymes). The affinity of M. for oxygen is higher than that of hemoglobin. Muscle has a high content of M. especially the cardiac muscle of diving mammals, such as whale and seal, which contains up to 8% M., compared with 0.5% M. in the cardiac muscle of dog. High levels of M. are also found in the flight muscles of birds. M. and hemoglobin were the first globular proteins submitted to structural elucidation by X-ray diffraction analysis. M. possesses no disulfide bridges or free SH-groups. It contains 8 variously sized right-handed helical regions, joined by non-ordered or random coil regions. 121 Amino acids

are involved in helix formation, representing a total  $\alpha$ -helix content of 77%. These 8 helices (A,B,C,D,E,F,G,H) are folded back on top of one another, and the heme is situated between helices E and F (see Proteins, Fig. 9). The molecule has the shape of a flattened ball with a pocket for the heme; the heme is almost totally buried, with just one edge exposed - that carrying the two hydrophilic propionic acid groups. The heme is held in position by a coordination complex between the central iron (II) atom and two histidine residues (on helices E and F, respectively). One of these histidines binds to the oxygen of the water molecule, which is bound to the heme. The position and the functional competence of the heme also depend upon the hydrophobic amino acids that line the inside of the heme pocket. Slight changes in the tertiary structure of M. destroy the oxygen binding function of the heme. Metmyoglobin, i.e. with Fe (III), does not bind oxygen. The primary and even tertiary structure of M. and hemoglobins from many different species show striking similarities. M. appears to be phylogenetically the oldest known heme protein, from which hemoglobin evolved as an independent molecule 600 million years ago. M. also appears to be phylogenetically related to Leghemoglobin (see).

**Myo-inositol:** a cyclitol found free and combined in animals and plants. The earlier equivocal name, *mesoinositol* should be avoided. M. is biosynthesized from D-glucose, and the configuration of C-atoms 1-5 is preserved during the conversion. M. is an essential yeast growth factor. It is present in relatively large amounts in brain cells, lens, thyroid gland, muscle, lung and liver. In the biosynthesis of M., a cyclase (EC 5.5.1.4) catalyses the conversion of glucose 6-phosphate into inositol 1-phosphate, which is dephosphorylated by the action of a phosphatase (EC 3.1.3.25). Inositol oxygenase converts M. into D-glucuronate, an intermediate in the Glucuronate pathway (see) of carbohydrate degradation.

M. occurs as its mono- and diphosphate esters in phospholipids and phosphoproteins. The M. phospholipids play a role in the response of mammalian cells to external stimuli such as hormones (see Inositol phosphates). M. hexaphosphate is called phytic acid: the mixture of Ca and Mg salts of phytic acid is called phytin. Phytic acid is an important phosphate storage compound in plant tissues, e.g. in cereal grains.



**Myo-inositol**

**Myokinase:** see Adenylate kinase.

**Myosin:** see Muscle proteins.

**Myrcene:** a triply unsaturated acyclic monoterpeno hydrocarbon. M. is a pleasant smelling liquid.  $M_r$  136.24, b.p.<sub>12</sub> 55-56 °C, p<sup>15</sup> 0.8013,  $n_D^{19}$  1.470. It is a component of many essential oils, and it is prepared for the perfumery industry by pyrolysis of  $\beta$ -

# Recombinant DNA

## A Short Course

James D. Watson

COLD SPRING HARBOR LABORATORY

John Tooze

EUROPEAN MOLECULAR BIOLOGY ORGANIZATION

David T. Kurtz

COLD SPRING HARBOR LABORATORY

SCIENTIFIC  
AMERICAN  
BOOKS

Distributed by

W. H. Freeman and Company

New York

Alternative Splicing Pathways Generate Different mRNAs from a Single Gene	97
Control Regions Are Found at the 5' and 3' Ends of Genes	98
Clustered Gene Families May Include Vestigial Evolutionary Relics	99
Gene Families May Be Expanded by Reverse Transcription of mRNA Molecules	100
Eukaryotic DNA Contains Interspersed Repetitive Sequences	100
Polypeptide Precursors Give Rise to Protein Hormones	101
<b>8. <i>In Vitro</i> Mutagenesis</b>	<b>106</b>
Deletions	106
Insertions	108
Substitutions: Deamination of Cytosine	109
Substitutions: Incorporation of Nucleotide Analogs	110
Substitutions: Misincorporation of Nucleotides	111
Mutants May Be Constructed by Using Oligonucleotides with Defined Sequences	111
<b>9. Rearranging Germ-Line DNA Segments to Form Antibody Genes</b>	<b>117</b>
The Basic Structure of Antibody Molecules Is Established	117
Separate Genes for V and C Segments Are Proposed	118
Messenger RNA Probes Are Used to Obtain Support for the Joining of V and C Genes	118
Functional Antibody Genes Are Isolated from Myeloma Cells	119
Embryonic Cells Are Sources of Unjoined V and C Genes	119
Multiple J (Joining) Segments Are Attached to Genomic C (Constant) Segments	119
Three Discontinuous Regions of DNA Code for Heavy-Chain Amino Acids	120
A DNA Elimination Event Allows a V <sub>H</sub> Gene to Be Attached to Two Different C <sub>H</sub> Genes	120
Alternative Splicing Allows Single Cells to Make Both $\mu$ - and $\delta$ -Class Heavy Chains with Identical V <sub>H</sub> Segments	121
Somatic Mutations Provide a Further Source of Immunoglobulin Diversity	122
Establishing the Genes of the Major Histocompatibility Complex (MHC) Proteins and Their Protein Antibodies Through Gene Cloning	123
<b>10. Tumor Viruses</b>	<b>127</b>
Cloning Integrated Forms of DNA Tumor Viruses	128
Tumor Proteins of SV40 and Polyoma	129

Cover illustration by Marvin Mattelson.

Library of Congress Cataloging in Publication Data  
Watson, James D., 1928–  
Recombinant DNA.

Includes bibliographies and index.

1. Recombinant DNA. I. Tooze, John. II. Kurtz, David T. III. Title. IV. Title: Recombinant D.N.A. [DNLM: 1. DNA, Recombinant. QU 58 W339r]  
QH442.W373 1983 574.87'3282 83-9069  
ISBN 0-7167-1483-3  
ISBN 0-7167-1484-1 (pbk.)

Copyright © 1983 by James D. Watson, John Tooze, and David T. Kurtz

No part of this book may be reproduced by any mechanical, photographic, or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted, or otherwise copied for public or private use, without permission from the publisher.

Printed in the United States of America

Scientific American Books is a subsidiary of Scientific American, Inc.

Distributed by W. H. Freeman and Company, 41 Madison Avenue, New York, New York 10010

2 3 4 5 6 7 8 9 0 KP 1 0 8 9 8 7 6 5 4

# 8

## *In Vitro* Mutagenesis

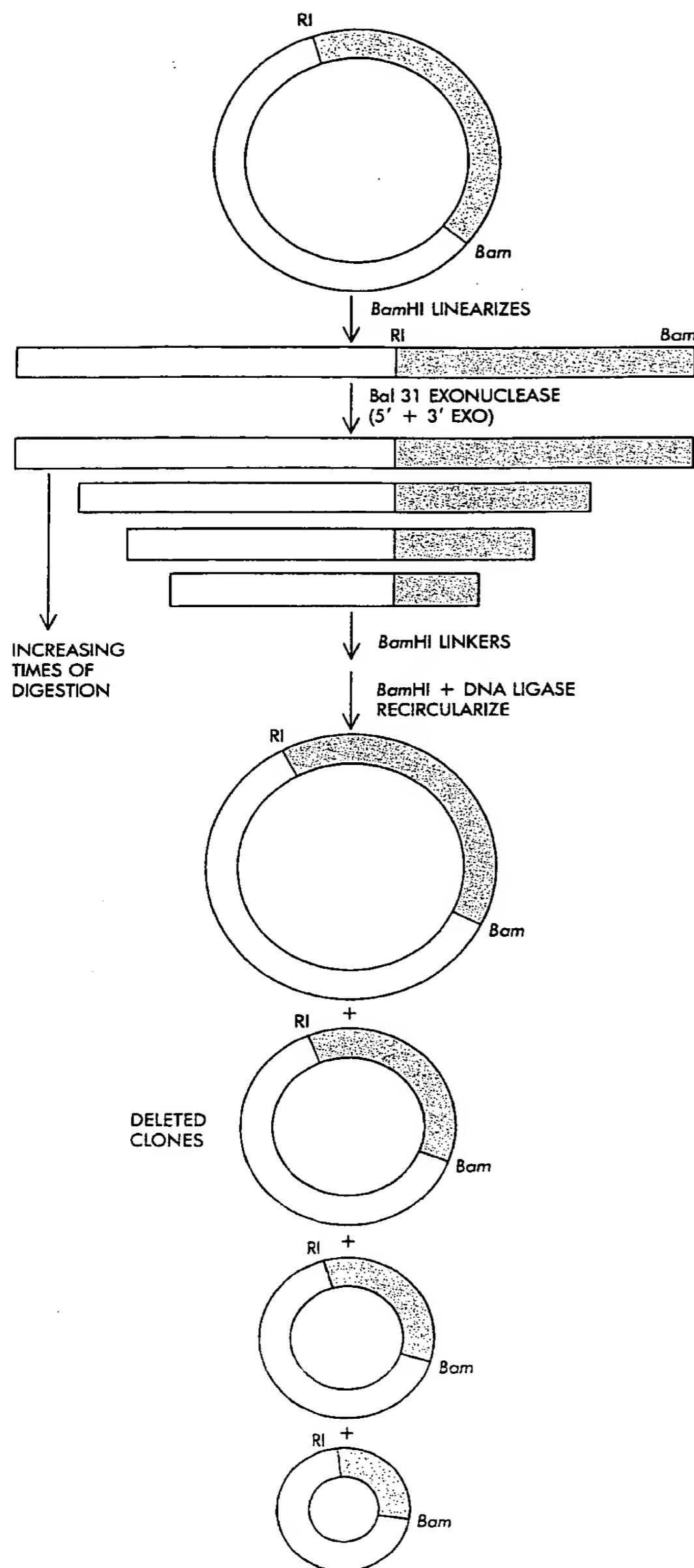
With the advent of recombinant DNA technology and ancillary techniques such as DNA sequencing, we can now examine, at the molecular level, the DNA sequences that are involved in the control of gene expression. The classical approach to genetics is to create *in vivo* mutations randomly throughout the genome, and then isolate those that display a particular phenotype. These mutants are then analyzed to determine which gene or genes have been altered. The precise nature of the mutation itself can be determined by DNA sequencing. An almost converse method (in fact, the method has been called "reverse genetics") is to create specific mutations in a DNA segment *in vitro*, and to analyze the effects of these changes on the organism *in vivo* following reintroduction of the mutant gene. It is now possible to create mutations—either deletions, insertions, or specific base changes—at predetermined sites in a DNA molecule.

### Deletions

The simplest *in vitro* mutation that can be constructed is the deletion of a DNA segment between two restriction enzyme sites. Circular SV40 DNA, for instance, can be partially digested with a given restriction enzyme that would normally produce several fragments. The conditions of digestion can be maintained so that each molecule gets, on the average, two or three cuts. The resultant molecules are recircularized using DNA ligase. The population will then consist of molecules

that have had a specific restriction enzyme fragment excised. Because deletions made in this way tend to be rather large, this approach is typically used in preliminary analysis to determine the functions of relatively large areas of a cloned DNA molecule. For instance, deletion of the SV40 *Hind* III B or C fragment destroys transforming activity, whereas deletion of the smallest *Hind* III fragment does not, but does prevent packaging of SV40 into virus. Once these relatively large areas of DNA have been associated with given functions, finer mutations can be made to determine more precisely the functional units of DNA.

Smaller deletions can be produced in a circular DNA molecule by cleaving it with a restriction enzyme that linearizes it (cleaves it once). The linear molecules are then treated with an enzyme called exonuclease III (exo III), which starts from each 3' end of the DNA and chews away single strands in a 3'-to-5' direction, creating a population of DNA molecules with single-stranded tails at each end. The tails can then be degraded with S1 nuclease, which specifically attacks single-stranded DNA, resulting in duplex DNA molecules with deletions. More recently, the enzyme Bal 31 exonuclease has been used; this enzyme chews away *both* strands from the ends of linear DNA molecules. These nucleolytic reactions can be controlled by varying the time of incubation, the temperature, and the enzyme concentration to make deletions that can range from 20 to 2000 bases of DNA. The deleted molecules are then



recircularized with DNA ligase (usually after addition of a synthetic oligonucleotide linker) and used to transform *E. coli* (Figure 8-1).

A variation of this technique was used to study the DNA sequences that are important in the regulation of transcription of the herpes simplex virus thymidine kinase gene (HSV tk). Two libraries of deletion mutants were constructed with the exo III and S1 nucleases: One set of mutations started from a site well beyond the 5' end and proceeded towards the gene; the other set started at a point in the gene and proceeded in the opposite direction through the 5' end of the gene.

DNA sequencing was used to define the precise endpoints of the two sets of deletions. Forty-three different mutants in which the deletions came from the 5' direction and 42 mutants in which the deletions came from the 3' direction were found to have terminated in a small 140-nucleotide stretch surrounding the 5' end of the tk gene. (The ends of all the deletions contained a synthetic *Bam* HI linker). Next, a search was made for deletion pairs that came from opposite directions and that terminated exactly 10 bases apart. When two such deletions were recombined through the *Bam* HI linker, the 10 bases of the linker replaced the 10 nucleotides in the normal sequence. This was called a "clustered point mutant" because most, though generally not all, of the 10 bases in the linker were different from the wild-type sequence (Figure 8-2, page 108). When two deletion mutants coming from opposite directions had their termini 15 or 20 bases apart, then recombination of these termini through the *Bam* HI linker resulted in a 5-or-10-base-pair deletion of DNA, as well as the base changes caused by the *Bam* HI linker. In this way, the spacing between control regions was systematically and precisely varied.

**Figure 8-1**  
Small deletions are made in a circular DNA molecule by cleaving it once with *Bam* HI and treating it with Bal 31 exonuclease, which digests the ends of both strands. The size of the deletion can be controlled by varying incubation time, temperature, and enzyme concentration. The strand is then recircularized by adding *Bam* HI linkers and DNA ligase.

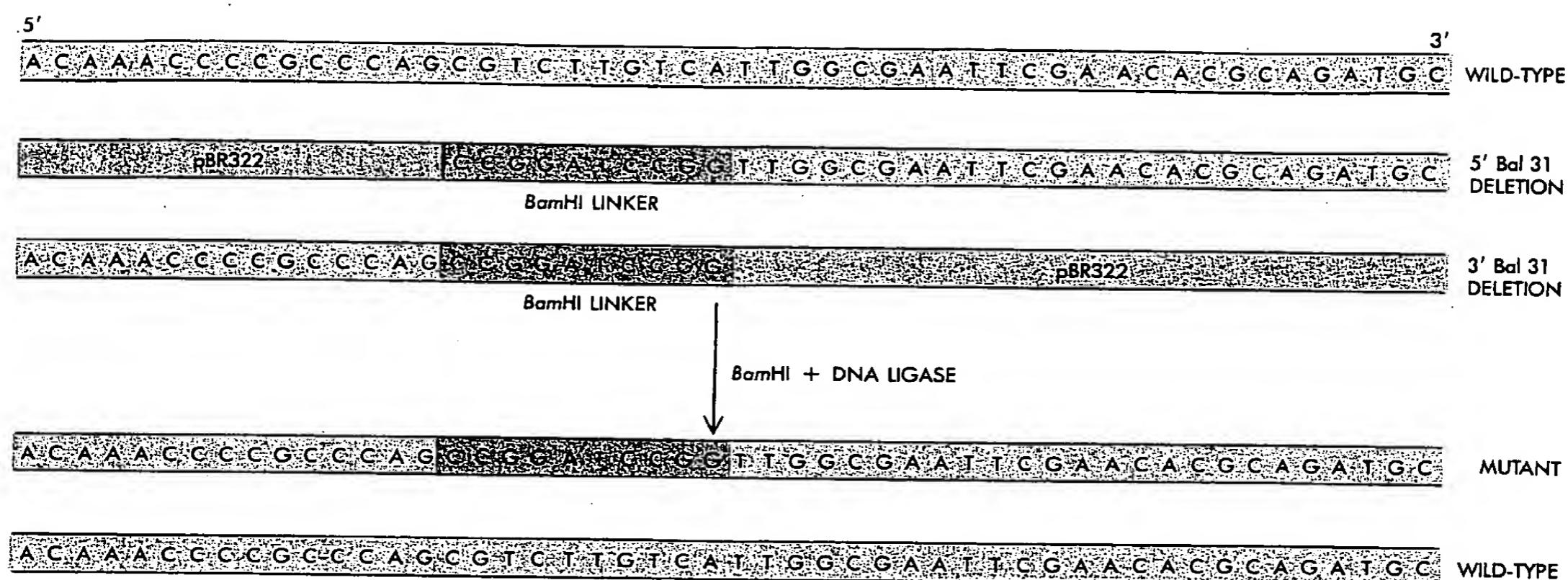


Figure 8-2

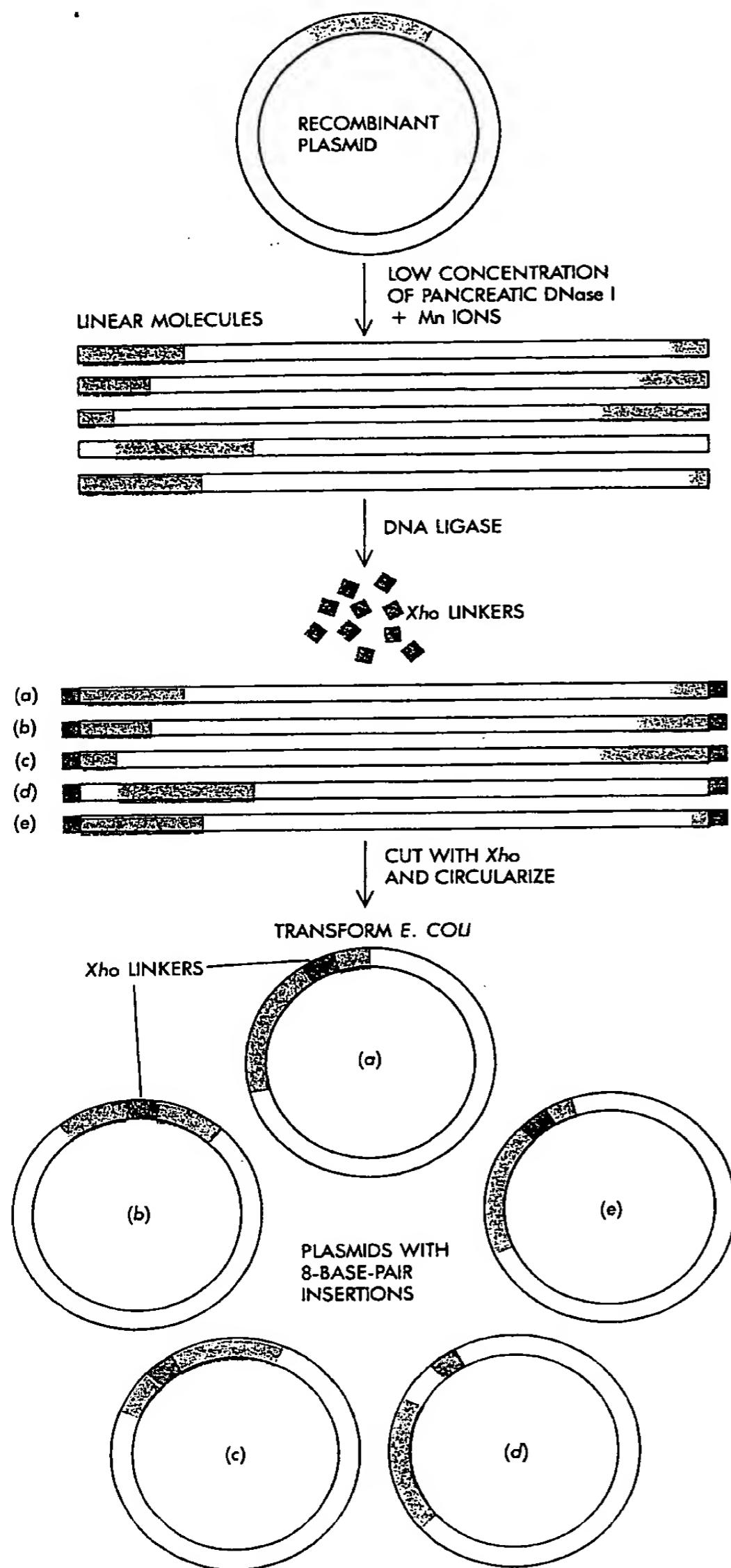
Creation of a clustered point mutation resulting in an eight-base substitution. Pairs of mutants in which the members terminate exactly ten bases from one another are selected. Identical *Bam*HI linkers are added to the ends; these ends are cleaved with *Bam*HI and joined with DNA ligase. A DNA molecule that differs from the wild-type sequence only in the ten bases that constituted the linkers is thus formed. Two of these bases happen to match the bases in the wild-type sequence; the other eight (boldface) represent the clustered point mutation.

A set of clustered point mutations or small deletions was constructed in this way throughout the promoter region of the HSV tk gene. When such mutants were tested in an *in vitro* transcription assay, it was found that three distinct regions are required for efficient expression of tk mRNA. One is the "TATA box" (Chapter 7, page 98). This AT-rich stretch of nucleotides about 25 to 30 bases upstream from the cap site is believed to direct initiation at the correct nucleotide. Two other regions near the tk gene were also found to be critical for transcription: GC-rich stretches at about -50 and -90. Base substitutions in either of these regions greatly reduce transcriptional efficiency. The spacing between these three regions was also found to be important: Deletions that bring the -90 region closer to the -50 reduce transcription, as do deletions that bring the -50 region closer to the TATA box.

#### Insertions

Insertion of a synthetic oligonucleotide linker at a given site in a cloned DNA molecule can be used

either to interrupt the normal DNA sequence or to generate a new restriction site that can serve as a starting point for further mutagenesis. Pancreatic DNase in the presence of manganese ions introduces random double-stranded breaks in a DNA molecule. Normally this enzyme would completely degrade a DNA molecule into small pieces; however, if cloned DNA is treated with a very small amount of pancreatic DNase for a very short time, each DNA molecule will be cut, on the average, only once. These linear molecules can be isolated from a gel. Synthetic oligonucleotide linkers can then be added to them so that they may be recircularized with DNA ligase. This will result in a population of DNA molecules with the linker sequence randomly inserted (Figure 8-3). Alternatively, cloned DNA molecules can be cleaved with a restriction enzyme that would normally cut each DNA molecule several times, but a very small amount is used so that each molecule is cut only once. Insertion of linkers into this population will result in a less random set of insertion mutants than that generated using pancreatic DNase, since each



**Figure 8-3**  
Creation of insertion mutants via synthetic oligonucleotide linkers. Recombinant plasmids are cleaved only once at random locations by treatment with a tiny amount of pancreatic DNase I. *Xho* linkers added to the resulting linear molecules are cut open with *Xho*. The molecules are recircularized with DNA ligase, producing plasmids that contain eight-base-pair inserts.

insertion will be at a known restriction enzyme site.

This technique was used to define DNA sequences important in mating-type regulation in yeast (Chapter 11). The yeast genome contains three copies ("cassettes") of genes that control mating type: one expressed copy (the "MAT" copy) and two silent copies (called "HMR" and "HML"). The silent genes are kept silent by the action of another gene set called "SIR." Insertion mutations around the HMR locus showed that the negative regulation of this gene by SIR is dependent upon two regions of DNA, one lying on each side of the silent cassette. Mutations in one of the regions (the "E," or "essential," region) resulted in complete loss of control by SIR, whereas mutations in the other (the "I," or "important," region) caused only a partial loss of regulation. Thus, a region of negative transcriptional control extending about 3000 bases was defined.

#### Substitutions: Deamination of Cytosine

Specific single base changes (point mutations) can also be generated in a DNA molecule *in vitro*. One method is based on the fact that cytosine residues in single-stranded, but not double-stranded, DNA can be chemically deaminated to produce uracil by the action of bisulfite ions. Thus, if small single-stranded regions can be created in a cloned DNA molecule, the C residues in this region can be specifically changed to U (which is read as T when the molecule is replicated).

There are two methods of creating single-stranded regions of DNA around a restriction enzyme site. One method is to use the enzyme exonuclease III to chew away single strands of DNA in the 3'-to-5' direction. The result will be single-stranded tails, the lengths of which will depend on the extent of the reaction with exo III. The other method, which is more controllable, is based on the fact that many DNA polymerases will, in the absence of dNTPs, degrade single strands of duplex DNA in the 3'-to-5' direction; if only one dNTP is present in the reaction, the polymerase will degrade single strands of duplex DNA until it reaches a point at which the base on the other strand is complementary to the available nucleotide. The exonucleolytic activity will then stop. Both phage T4 DNA polymerase and the so-called

"Klenow" or large fragment of *E. coli* DNA polymerase I have this property. For example, if a DNA molecule is cut with the restriction enzyme *Sma*I, which recognizes the sequence CCCGGG, and this cleaved molecule is treated with Klenow polymerase in the presence of only dATP, the enzyme starts at the *Sma*I site and chews away each 3' end of the duplex DNA until it reaches an A residue. This is usually only a few bases away from the enzyme site; thus small single-stranded regions are generated on each strand. The molecule is then treated with bisulfite, which changes the C residues to U. Next Klenow polymerase is added in the presence of *all four* dNTPs and the single-stranded regions are repaired; the uracil residues created by the bisulfite pair with A. GC base pairs are thus mutated to AT pairs (Figure 8-4).

Using this method, a number of single and double base changes were generated in a cloned proline tRNA molecule. The mutations were GC → AT changes clustered around a *Sma*I site inside the tRNA molecule. When tested in *in vitro* transcription assays, some of these mutants were found to be very poorly transcribed. This indicated that the promoter for the tRNA genes (which are transcribed by RNA polymerase III) is actually *inside* the gene, rather than off the 5' end. This "intragenic promoter" has been found for all

genes transcribed by RNA polymerase III, in contrast to genes transcribed by RNA polymerase I or II or by the prokaryotic RNA polymerases.

#### Substitutions: Incorporation of Nucleotide Analogs

Another method of producing single base changes near a restriction site is to create small single-stranded regions in a cloned DNA molecule. The molecule is treated with a restriction enzyme in the presence of ethidium bromide. This drug causes the restriction enzyme to "nick" the DNA at the restriction site (that is, to cut only one of the DNA strands rather than both). This nick can be extended into a small gap by using a low concentration of DNA polymerase isolated from a bacterium called *M. luteus*. In the absence of triphosphates, *M. luteus* polymerase will degrade only five or six bases of DNA in the 5'-to-3' direction, starting from the nick. The result is a DNA molecule with a five-or-six-base gap in it. The gap is then repaired with Klenow polymerase in the presence of dATP, dCTP, dGTP, and N-4-hydroxycytosine instead of dTTP. This nucleotide analog is incorporated in place of T, and, because the 6-keto-enol ratio ( $C_6=O/C_6OH$ ) of this compound is almost 1, it can pair equally well with

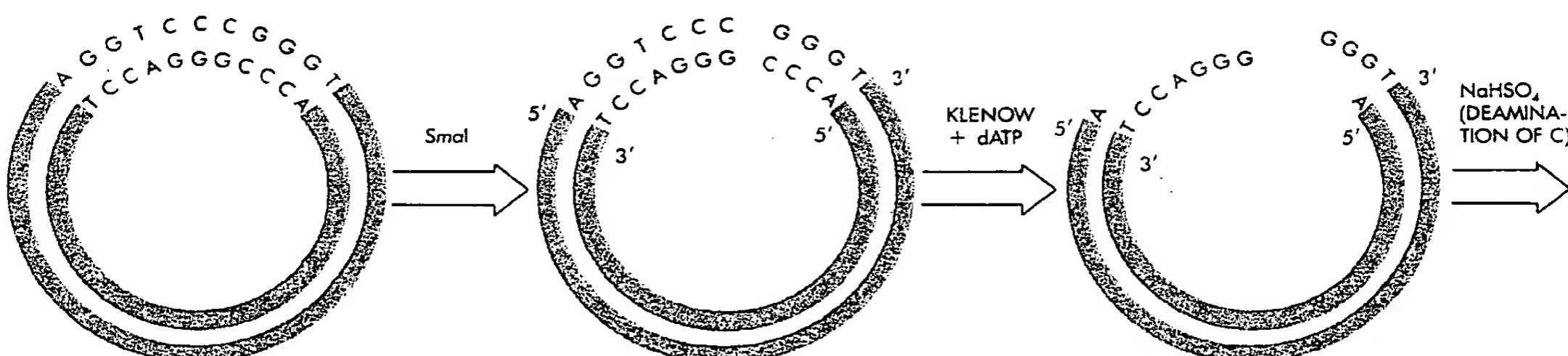


Figure 8-4

Creation of a substitution mutant through deamination of cytosine. A *Sma*I cleavage site is treated with Klenow polymerase in the presence of dATP. The enzyme chews away the single strands in the 3'-to-5' direction, until it reaches an A residue on each strand. The addition of bisulfite changes the exposed C residues to U. Klenow polymerase is added again, this time with all four dNTPs; the result is a repaired molecule in which two of the GC base pairs have been mutated to AU (or AT).

either A or G. Incorporation of a G residue in the complementary strand will result in an  $AT \rightarrow GC$  mutation (Figure 8-5, page 112).

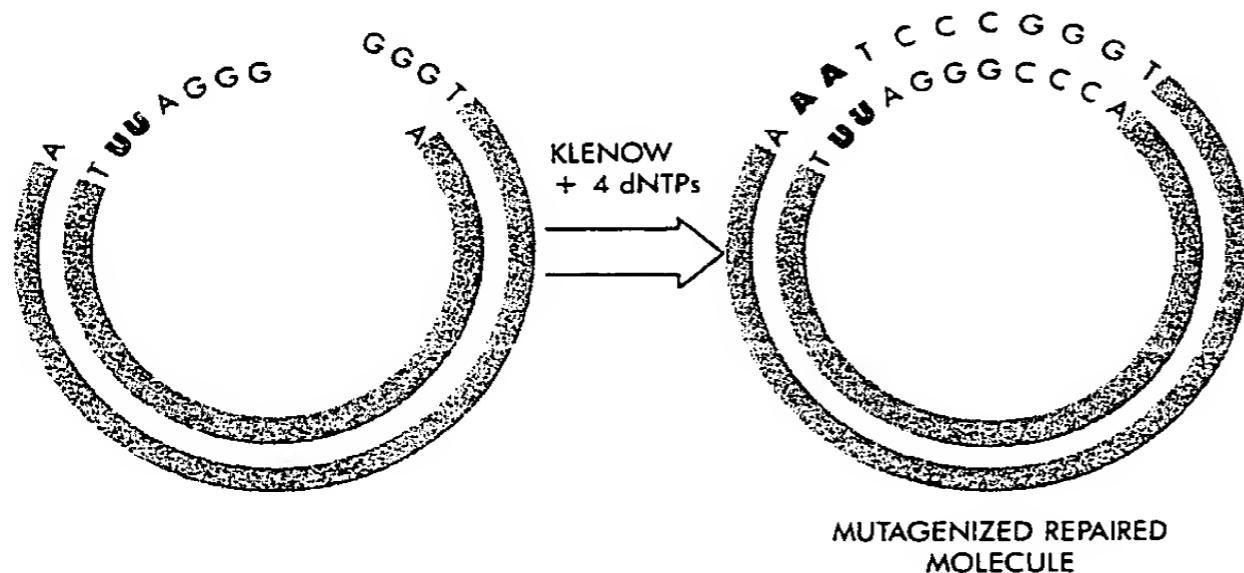
#### Substitutions: Misincorporation of Nucleotides

Finally, single base changes near a restriction site in DNA can be generated by misincorporation of a nucleotide during enzymatic repair of a gap, created as just described. The gapped molecules are treated with *E. coli* or T4 DNA polymerase in the presence of only three of the four dNTPs. Normally, the polymerase stops when it reaches a point at which it has no complementary nucleotide for the base in the opposite strand; a small percentage of the time, however, the polymerase will misincorporate one of the other three nucleotides in place of the one that is missing, and then continue to repair the gap completely. This results in circular plasmids with single base mismatches. When the plasmids are reintroduced into *E. coli*, the mismatches will be repaired during the next round of DNA replication. Half of the replicating strands will recreate the wild-type sequences, and the other half will create mutant sequences with a single base-pair change. Since the base change was made at the restriction site at which the molecule

was originally nicked, the site has been altered and the mutant molecule will not be cleaved by that enzyme (Figure 8-6, page 113). The daughter plasmid molecules can be quickly screened for the presence or absence of the enzyme site to determine whether they are mutant or wild-type. One potential problem with this technique is the fact that most polymerases have a "proofreading" function, in the form of a nuclease that excises a mismatched nucleotide as soon as it is incorporated. This can be circumvented by using  $\alpha$ -thiophosphate nucleotides during the gap repair; these molecules cannot be excised by the polymerase's proofreading exonuclease.

#### Mutants May Be Constructed by Using Oligonucleotides with Defined Sequences

All of the techniques we have described for generating specific base changes in a DNA molecule are dependent upon the presence of a restriction enzyme site near the region of interest. Of course, the location of such a site is a matter of random chance. Another method for making specific point mutants in a cloned DNA molecule does not depend at all on the presence of a convenient restriction site. The method utilizes synthetic oligonucleotides of defined sequence (Chapter 5, page 63). When the sequence of a cloned DNA mole-



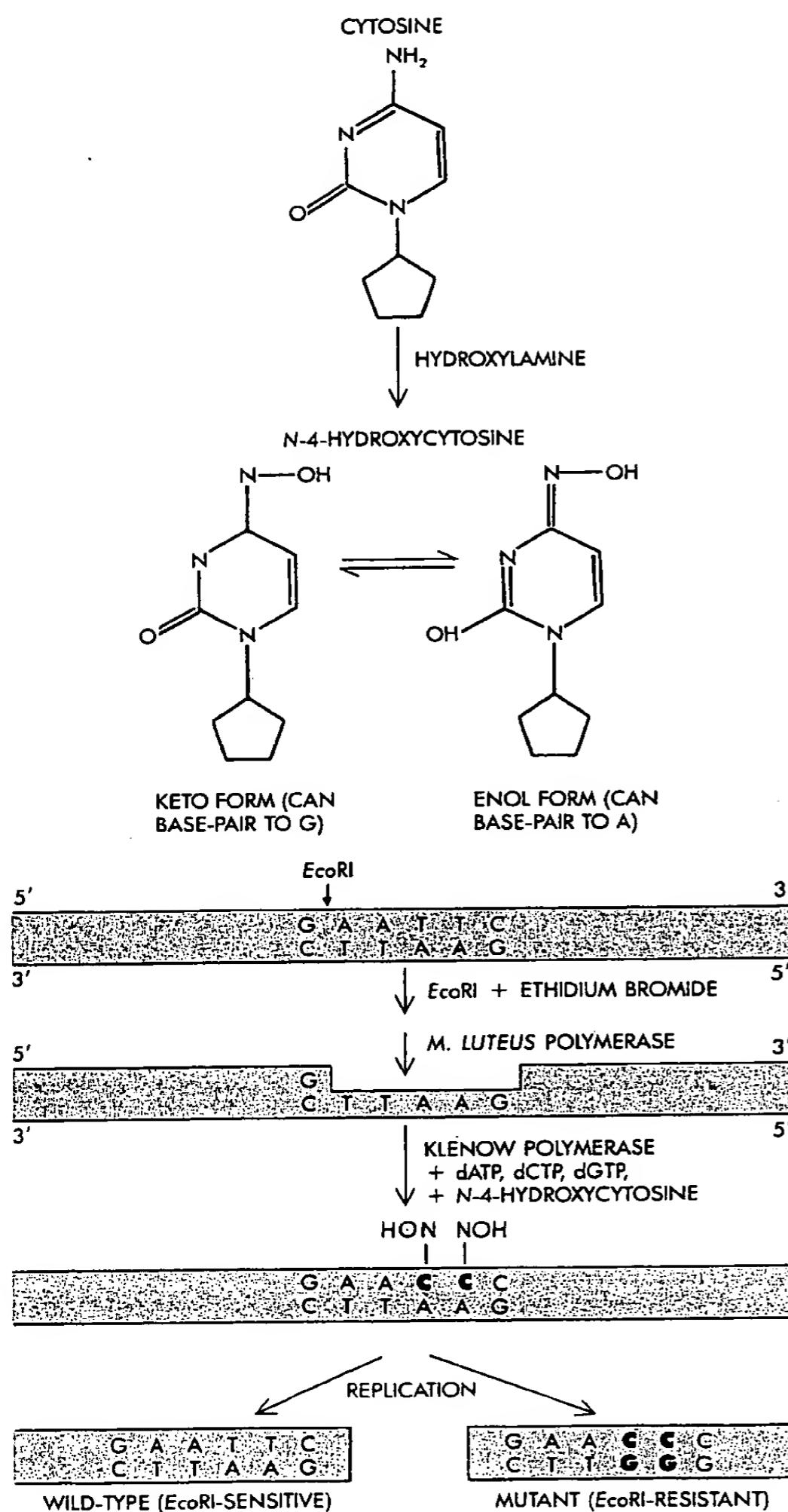


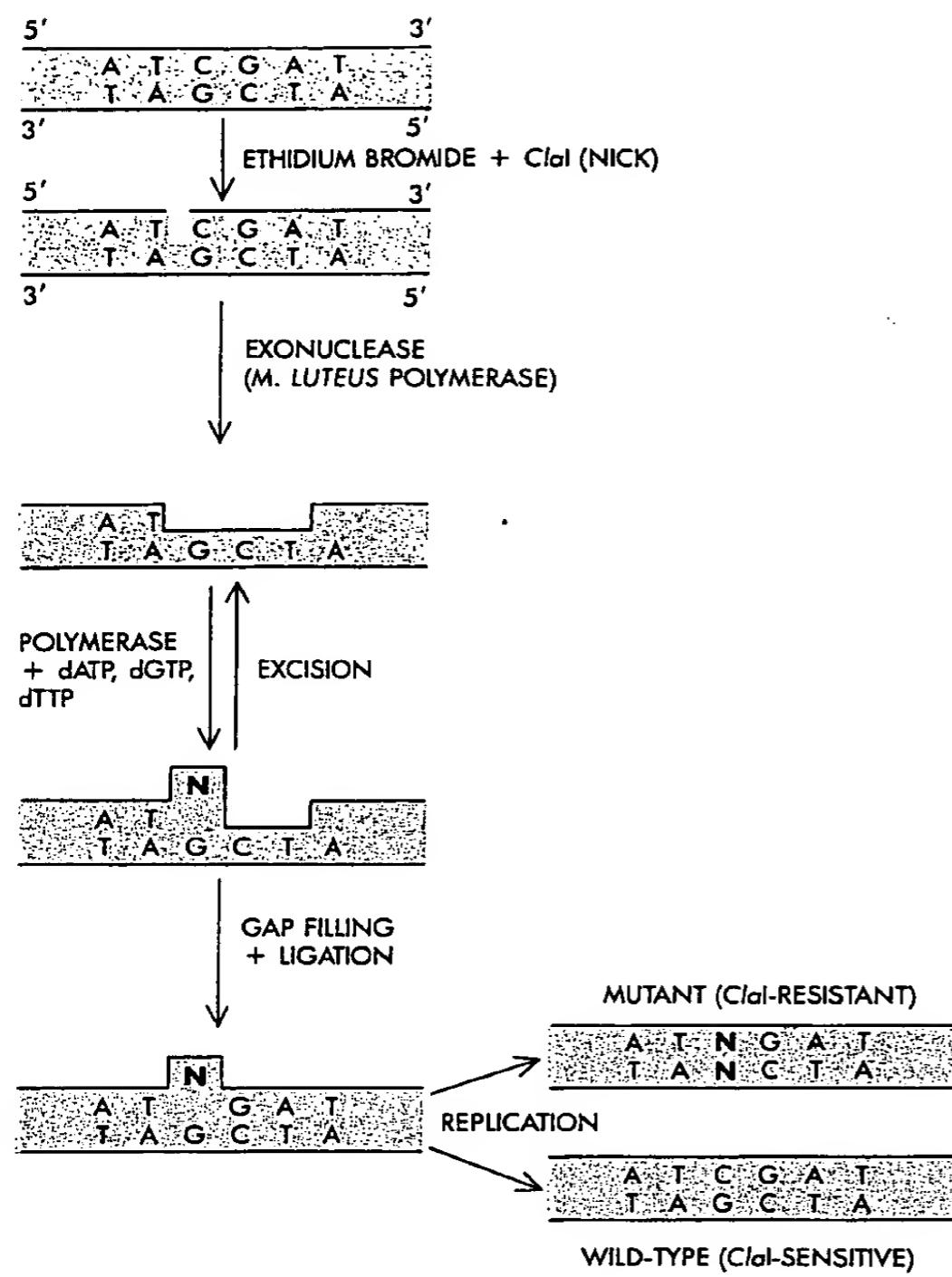
Figure 8-5

*N*-4-Hydroxycytosine can be obtained from cytosine by treatment with hydroxylamine. *N*-4-Hydroxycytosine can pair equally well with A or G. A DNA strand is nicked at an *Eco* RI site by cleaving it with this restriction enzyme in the presence of ethidium bromide. The nick can be extended into a small gap by using *M. luteus* polymerase. The gap is then repaired by treating it with Klenow polymerase in the presence of dATP, dCTP, dGTP, and *N*-4-hydroxycytosine. The latter replaces T in the resulting double strand, since in its enol form it can pair with A. During replication, the *N*-4-hydroxycytosine residues are interpreted as C residues, so the complementary strand is formed with G residues opposite the inserts, instead of the original A residues. The result is a mutant that resists *Eco* RI cleavage because the recognition site for that enzyme has been altered.

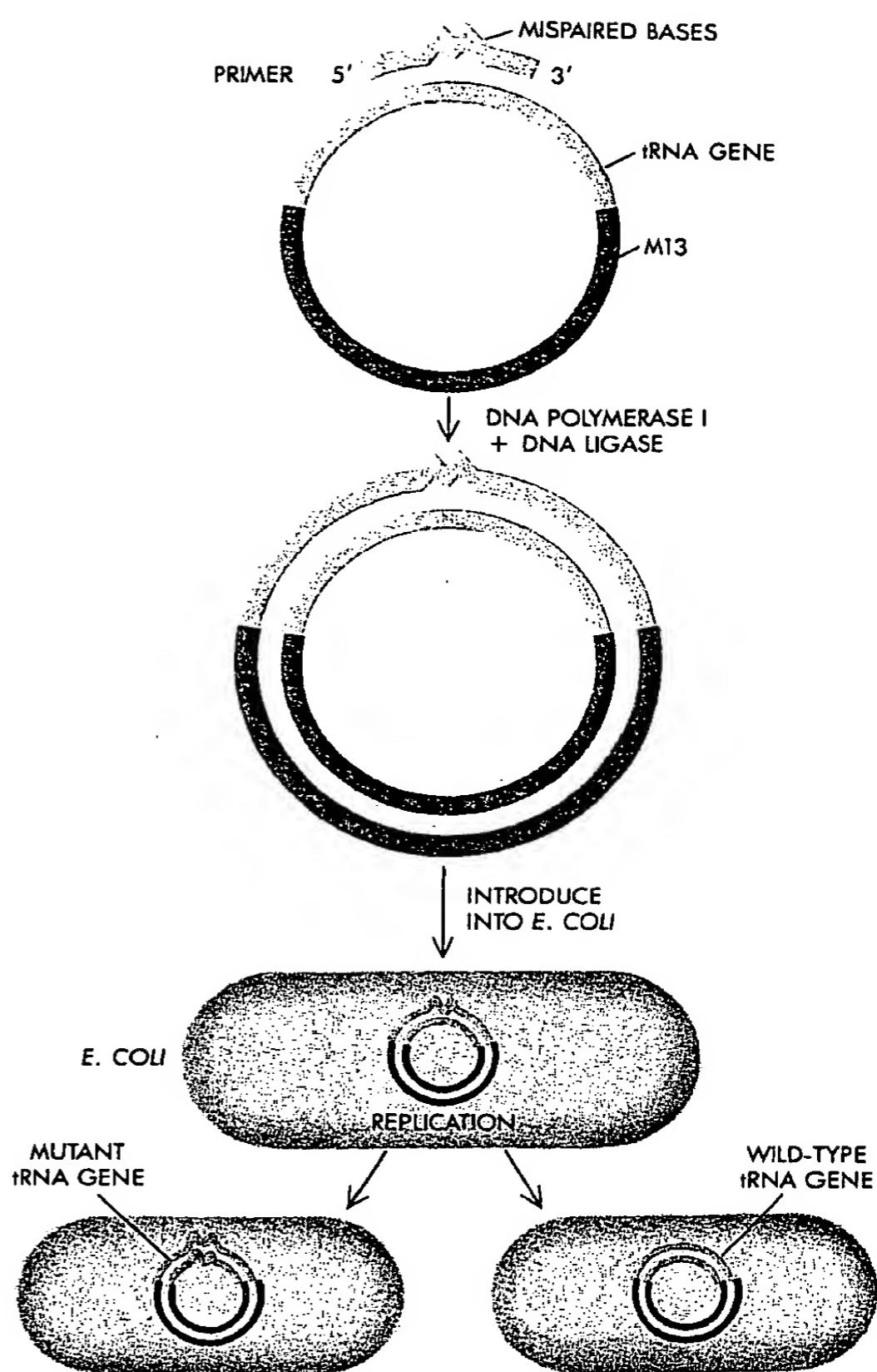
matches are in the middle of the oligonucleotide rather than at an end. The mismatched oligonucleotide serves as a primer for the action of DNA polymerase to synthesize the remainder of the complementary strand. The double-stranded molecule, now containing one or two mismatches, is introduced into *E. coli*, where the mismatches will be repaired, to recreate either the wild-type sequence or the mutant one (Figure 8-7).

Theoretically, 50 percent of the daughter molecules will be wild-type and 50 percent mutant; in practice, however, the percentage of mutant molecules is much lower (usually 10 to 15 percent). [This is probably for one or both of the following reasons: (1) The DNA polymerase, having synthesized the entire complementary strand starting from the primer, sometimes continues polymerizing after going around the circle, so that it actually displaces the primer and recreates the original wild-type sequence. (2) The original wild-type DNA strand is "marked" somehow, perhaps by methylation, in such a way that when the mismatched duplex is introduced into *E. coli*, it is repaired to the wild-type sequence more often than to the mutant sequence.] Mutant molecules can be distinguished from wild-type ones in two ways. If the base change either created or destroyed a restriction enzyme site, then several M13 clones can be quickly assayed for the presence or absence of that site. Alternatively, the oligonucleotide that was originally used to make the mutation can be used to distinguish mutant from wild-type

cule has been determined, this information is used to make an oligonucleotide that is 12 to 15 bases long and that is complementary to the region to be mutated, but with one or two mismatches. This oligonucleotide is mixed with a single-stranded clone of the complementary strand of the original molecule, carried in an M13 phage vector. Although the oligonucleotide is not a perfect match, it will anneal to the single-stranded clone if the hybridization conditions are not very stringent (the annealing must be done at a low temperature and in the presence of high salt), and if the mis-



**Figure 8-6**  
Creation of a substitution mutant through misincorporation of a nucleotide. The restriction enzyme *Cla*I is used in the presence of ethidium bromide to nick a cloned DNA molecule. The nick is extended into a small gap by using *M. luteus* polymerase. The gap is then repaired by treating it with *E. coli* or T4 DNA polymerase and three of the four dNTPs—dCTP is omitted. Occasionally the enzyme will try to substitute A, T, or G opposite the G residue in the intact chain. If the mismatched nucleotide is not excised, the result will eventually be a set of daughter plasmid molecules that contain a mutant base pair.

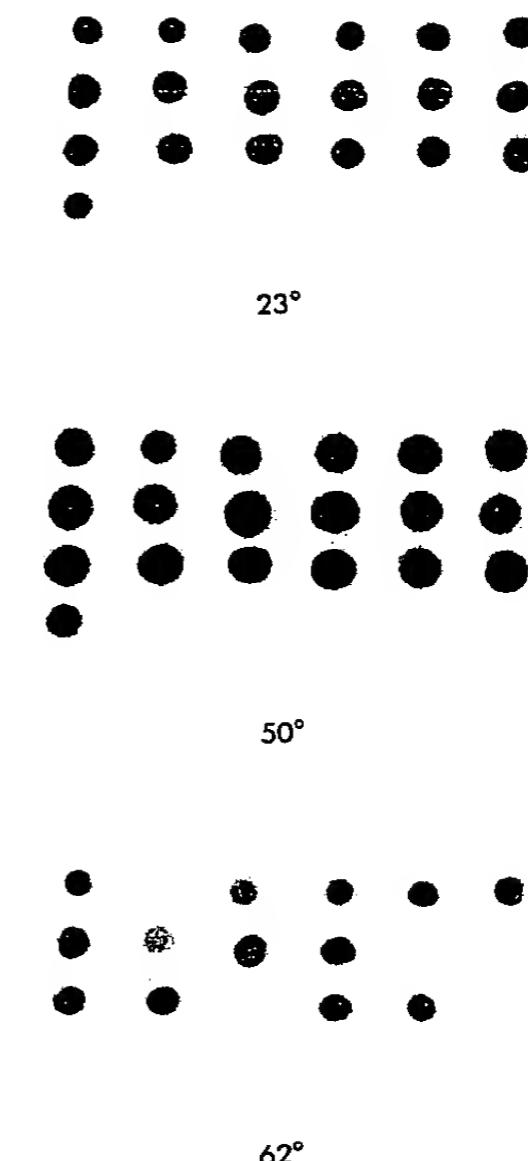


**Figure 8-7**  
Creation of a substitution mutant by use of a synthetic oligonucleotide. The 12-to-15-base oligonucleotide is constructed so that it is complementary to a region of a DNA strand, but with one or two mismatches. When mixed with a clone of the complementary strand, the oligonucleotide will anneal to it even though the match is not exact, as long as the hybridization conditions are not stringent and the mismatches are in the middle of the oligonucleotide segment. The segment then serves as a primer for DNA polymerase I, which synthesizes the remainder of the complementary strand. When the resulting double-stranded molecule is introduced into *E. coli*, the molecule replicates to recreate either the original wild-type sequence or the mutant sequence.

molecules: As was described above, an oligonucleotide will hybridize to a complementary sequence with one or two mismatches if the stringency of hybridization is kept low; if the temperature of the hybridization reaction is raised, however, the oligonucleotide will form a stable duplex only with a sequence to which it is perfectly complementary. The oligonucleotide that was used to make the mutation can be labeled with  $^{32}\text{P}$  and used as a probe to screen bacterial colonies on nitrocellulose filters, as described earlier (Figure 6-7). If the temperature of the hybridization is raised in  $5^\circ\text{C}$  increments, a point can usually be reached at which the labeled oligonucleotide will hybridize only to the mutant molecules (to which it is perfectly complementary) and not to the wild-type molecules (Figure 8-8).

A synthetic oligonucleotide was used to create a synthetic suppressor tRNA. A 12-base oligonucleotide was synthesized to be complementary to the anticodon loop of a cloned lysine tRNA molecule, except that the bases complementary to the lysine anticodon AAA were changed to be complementary to the termination codon UAG. This mismatched oligonucleotide was used to generate a mutant lysine tRNA molecule. The mutant tRNA was injected into frog oocytes with a cloned mutant  $\beta$ -globin gene obtained from a thalassemia patient (Chapter 17). This globin gene had the termination codon UAG at amino acid position 17, instead of the codon AAG which codes for lysine. Normally, the synthesis of  $\beta$ -globin polypeptide in oocytes would stop at this UAG codon. When the mutant tRNA gene is also microinjected, however, lysine was incorporated at this position, and synthesis of the  $\beta$ -globin protein proceeded normally.

Oligonucleotide-directed mutagenesis is also proving valuable for the study of protein structure and function. If the DNA sequence of the protein-coding region of a gene is known, then a synthetic oligonucleotide can be used to specifically change one amino acid codon to another. When this mutant gene is reintroduced into cells by means of techniques that we will describe later (Chapter 14), it will produce a protein with exactly one amino acid change. The effects of such an alteration on the protein's function can then be assessed.



**Figure 8-8**  
Distinguishing mutants created by defined-sequence oligonucleotides. M13 clones, some of which contain a specific base change, were obtained as described in Figure 8-7. DNA from these clones was spotted onto a nitrocellulose filter. The oligonucleotide that had been used to create the mutation was labeled with  $^{32}\text{P}$  and hybridized to the filter. The filter was then washed at successively higher temperatures until the M13 clones containing the mutant sequence could be distinguished from those harboring the wild type. (Courtesy of Mark Zoller.)

This technique is currently being used to study the amino acids that are responsible for the correct insertion and anchorage of membrane proteins, and also for the study of the structure of certain tumor virus proteins that may be involved in oncogenic transformation.

Given the recent advances in DNA sequencing and in oligonucleotide synthesis, oligonucleotide-directed mutagenesis could prove to be by far the most versatile method of creating specific point mutations in a DNA molecule.

## READING LIST

## Original Research Papers (Reviews)

## DELETIONS

Lai, C. J., and D. Nathans. "Deletion mutants of SV40 generated by enzymatic excision of DNA segments from the viral genome." *J. Mol. Biol.*, 89: 179-193 (1974).

Mertz, J. E., J. Carbon, M. Herzberg, R. W. Davis, and P. Berg. "Isolation and characterization of individual clones of simian virus 40 mutants containing deletions, duplications and insertions in their DNA." *Cold Spring Harbor Symp. Quant. Biol.*, 39: 69-84 (1974).

Beier, D. R., and E. T. Young. "Characterization of a regulatory region upstream of the ADR 2 locus of *S. cerevisiae*." *Nature*, 300: 724-728 (1982).

Grummt, I. "Nucleotide sequence requirements for specific initiation of transcription by RNA polymerase I." *Proc. Natl. Acad. Sci. USA*, 79: 6908-6911 (1982).

McKnight, S. L., and R. Kingsbury. "Transcriptional control signals of a eukaryotic protein-coding gene." *Science*, 217: 316-324 (1982).

Mirault, M. E., R. Southgate, and E. Delwart. "Regulation of heat shock genes: A DNA sequence upstream of *Drosophila hsp* genes is essential for induction." *EMBO J.*, 1: 1279-1282 (1982).

Struhl, K. "The yeast *his3* promoter contains at least two distinct elements." *Proc. Natl. Acad. Sci. USA*, 79: 7385-7389 (1982).

## INSERTIONS

Heffron, F., M. So, and B. J. McCarthy. "In vitro mutagenesis of a circular DNA molecule using synthetic restriction sites." *Proc. Natl. Acad. Sci. USA*, 75: 6012-6016 (1978).

Abraham, J., J. Feldman, K. A. Nasmyth, J. N. Strathern, A. J. S. Klar, J. R. Broach, and J. B. Hicks. "Sites required for position-effect regulation of mating type information in yeast." *Cold Spring Harbor Symp. Quant. Biol.*, 47: 989-998 (1983).

## BASE CHANGES

Chu, C. T., D. S. Parris, R. A. F. Dixon, F. E. Farber, and P. A. Schaffer. "Hydroxylamine mutagenesis of HSV DNA and DNA fragments: Introduction of mutations into selected regions of the viral genome." *Virology*, 98: 168-181 (1979).

Shortle, D., and D. Nathans. "Regulatory mutants of simian virus 40: Constructed mutants with base substitutions at the origin of viral DNA replication." *J. Mol. Biol.*, 131: 801-817 (1979).

Shortle, D., K. Koshland, G. M. Weinstock, and D. Botstein. "Segment-directed mutagenesis: Construction in vitro of point mutations limited to a small predetermined region of a circular DNA molecule." *Proc. Natl. Acad. Sci. USA*, 77: 5375-5379 (1980).

Hayatsu, H. "Bisulfite modification of nucleic acids and their constituents." *Prog. Nuc. Acid Res. Mol. Biol.*, 16: 75-124 (1976).

Shortle, D., and D. Botstein. "Single-stranded gaps as localized targets for in vitro mutagenesis." In J. F. Lemont and W. M. Generosa, eds., *Molecular and Cellular Mechanisms of Mutagenesis*. Plenum, New York, 1982, pp. 147-156. (Review.)

Weiher, H., and H. Schaller. "Segment-specific mutagenesis: Extensive mutagenesis of a *lac* promoter/operator element." *Proc. Natl. Acad. Sci. USA*, 79: 1408-1412 (1982).

Ciampi, M. S., D. A. Melton, and R. Cortese. "Site-directed mutagenesis of a tRNA gene: Base alterations in the coding region affect transcription." *Proc. Natl. Acad. Sci. USA*, 79: 1388-1392 (1982).

Sakonju, S., D. Bogenhagen, and D. D. Brown. "A control region in the center of the 5S gene directs specific initiation of transcription, I: The 5' border of the region." *Cell*, 19: 13-26 (1980).

Bogenhagen, D., S. Sakonju, and D. D. Brown. "A control region in the center of the 5S gene directs specific initiation of transcription, II: The 3' border of the region." *Cell*, 19: 27-36 (1980).

Shortle, D., J. Pipas, S. Lazarowitz, D. DiMaio, and D. Nathans. "Constructed mutants of simian virus 40." In J. K. Setlow and A. Hollaender, eds., *Genetic Engineering, Principles and Methods*, vol. 1. Plenum, New York, 1979, pp. 73-92. (Review.)

Hochstadt, J., H. L. Ozer, and C. Shopsis. "Genetic alteration in animal cells in culture." *Curr. Top. Microbiol. Immunol.*, 94/95: 244-308 (1981). (Review.)

Everett, R. D., and P. Chambon. "A rapid and efficient method for region- and strand-specific mutagenesis of cloned DNA." *EMBO J.*, 1: 433-437 (1982).

Hirose, S., K. Takeuchi, and Y. Suzuki. "In vitro characterization of the fibroin gene promoter by the use of single-base substitution mutants." *Proc. Natl. Acad. Sci. USA*, 79: 7258-7262 (1982).

Peden, K. W., and D. Nathans. "Local mutagenesis within deletion loops of DNA heteroduplexes." *Proc. Natl. Acad. Sci. USA*, 79: 7214-7217 (1982).

Zakour, R. A., and L. A. Loeb. "Site-specific mutagenesis by error-directed DNA synthesis." *Nature*, 295: 708-710 (1982).

#### MUTANT CONSTRUCTION BY USING OLIGONUCLEOTIDES WITH DEFINED SEQUENCES

Gillam, S., C. R. Astell, and M. Smith. "Site specific mutagenesis using oligodeoxyribonucleotides: Isolation of a phenotypically silent  $\phi$ X174 mutant, with a specific nucleotide deletion, at very high frequency." *Gene*, 12: 129-137 (1980).

Smith, M., and S. Gillam. "Constructed mutants using synthetic oligodeoxyribonucleotides as site-specific mutagens." In J. K. Setlow and A. Hollaender, eds., *Genetic Engineering, Principles and Methods*, vol. 3. Plenum, New York, 1981, pp. 1-32. (Review.)

Shortle, D., D. DiMaio, and D. Nathans. "Directed mutagenesis." *Ann. Rev. Genetics*, 15: 265-294 (1981).

Wallace, R. B., M. Schold, M. J. Johnson, P. Dembek,

and K. Itakura. "Oligonucleotide directed mutagenesis of the human  $\beta$ -globin gene: A general method for producing specific point mutations in cloned DNA." *Nuc. Acids Res.*, 9: 3642-3656 (1981).

Dalhadie-McFarland, G., L. W. Cohen, A. D. Riggs, C. Morin, K. Itakura, and J. H. Richards. "Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function." *Proc. Natl. Acad. Sci. USA*, 79: 6409-6413 (1982).

Laski, F. A., R. Belagaje, U. L. RajBhandary, and P. Sharp. "An amber suppressor tRNA derived by site-specific mutagenesis." *Proc. Natl. Acad. Sci. USA*, 79: 5813-5817 (1982).

Temple, G. F., A. M. Dozy, K. L. Roy, and Y. W. Kan. "Construction of a functional human suppressor tRNA gene." *Nature*, 296: 537-540 (1982).

Winter, G., A. R. Fersht, A. J. Wilkinson, M. Zoller, and M. Smith. "Redesigning enzyme structure by site-directed mutagenesis." *Nature*, 296: 756-758 (1982).